Accurate Molecular Electrostatic Potentials Based on Modified PRDDO/M Wave Functions. I. Electrostatic Potential Derived Atomic Charges

DENNIS S. MARYNICK

Valerian Software, 3058 Creekview Drive, Grapevine, Texas 76051

Received 13 June 1996; revised 11 October 1996

ABSTRACT: A new approach for the calculation of electrostatic potential derived atomic charges is presented. Based on molecular orbital calculations in the PRDDO/M approximation, the new parametrized electrostatic potential (PESP) method is parametrized against *ab initio* MP2/6−31G** calculations. For a data set of 820 atoms in 145 molecules containing H, C, N, O, F, P, S, Cl, and Br (including hypervalent species), the PESP method achieves a mean absolute error of 0.037 e^- with a correlation coefficient of 0.990. Unlike other approximate approaches, no scaling factor is required to improve the agreement between PESP charges and the underlying *ab initio* results. PESP calculations are an order of magnitude faster than the simplest *ab initio* calculation (STO-3G) on large molecules while achieving a level of accuracy that rivals much more elaborate *ab initio* methods. © 1997 by John Wiley & Sons, Inc. *J Comput Chem* **18**: 955–969, 1997

Keywords: electrostatic; potential; charge; PRDDO; PESP

Introduction

The molecular electrostatic potential (ESP) is a rigorously defined, real physical property that can be calculated directly from the charge distribution or molecular wave function.[1–3] In terms of the wave function, the ESP at the point r is defined as

$$V(r) = \sum_A \frac{Z_A}{|R_A - r|} - \int \frac{\psi^*(r')\psi(r')dr'}{|r' - r|} \quad (1)$$

Correspondence to: D. S. Marynick; e-mail: valerian@ix.netcom.com

In the self-consistent field, molecular orbital, linear combination of atomic orbital (SCF-MO-LCAO) approximation, this expression becomes

$$V(r) = \sum_A \frac{Z_A}{|R_A - r|} - \sum_{i,j} P_{ij} E_{ij}. \quad (2)$$

Here P_{ij} is a density matrix element and E_{ij} is the electronic part of the ESP operator expressed in terms of the AO basis functions ρ,

$$E_{ij} = \int \frac{\rho_i^* \rho_j}{|r - r'|} dr'. \quad (3)$$

CCC 0192-8651 / 97 / 070955-15

It is now well recognized[4–8] that ESPs play an important role in the analysis and understanding of long-range noncovalent interactions such as hydrogen bonding, solvation, crystal packing, and protein–ligand interactions. Another area in which ESPs play a key role, and the focus of the current work, is in the definition of atomic charges.[9–14] Because these charges are derived directly from the wave function and may be interpreted as the atom centered monopole approximation to the change density, they represent one of the most unambiguous and rigorous definitions of atomic charge and have generally been considered to be highly suitable for describing the long-range Coulombic interactions in molecular force fields.

ESP-derived atomic charges are obtained by a two-step process. First, the ESP is calculated from the wave function for a large number of points in the region outside the van der Waals envelope of the molecule. Second, a set of least squares equations are solved that define atomic charges that best reproduce the quantum mechanically derived ESPs subject to various constraints. One constraint that is almost always employed is that the atomic charges sum to the overall molecular charge. However, other conditions, such as constraining the dipole moment derived from the atomic charges to be equal to a specific value, are easy to impose. While a number of algorithms exist, the most common procedure employs the Connolly[15] algorithm for the selection of points. This procedure distributes points on atom-centered spheres with radii defined as specific multiples (greater than 1.0) of the van der Walls radii. Points inside the scaled atomic spheres, which may lie too close for the monopole approximation to be valid, are then eliminated. Typically, four spheres are employed, with radii 1.4, 1.6, 1.8, and 2.0 times the van der Waals radii.[11] The density of points is generally taken to be 1.0 point/Å, and the charges are derived via a noniterative least squares procedure that employs Lagrange multipliers for the constraints.[12, 16]

While the calculation of ESP-derived atomic charges is in principle straightforward, in practice it is not. Most workers in the field seem to accept charges derived from Hartree–Fock (HF) 6–31G* calculations as the *de facto* standard, even though it is clear that the HF wave function should in general be too ionic. Even at the aforementioned level of theory, the calculation of wave functions for typical molecules of biological interest (with 50 or more atoms) is not a trivial task. For this reason, a significant amount of work has appeared on the application of semiempirical methods to this problem. Two particularly useful articles in this area are that of Besler—Merz—Kollman[16] (BMK) and Merz.[17] BMK employed a set of 20 molecules containing 67 symmetry unique H, C, N, O, and S atoms and calculated ESP-derived charges from both MNDO and AM1 wave functions by explicitly deorthogonalizing the semiempirical wave function. They found the best agreement with HF/6–31G* charges using the MNDO Hamiltonian, obtaining an average absolute error of 0.125 e^-, an average scaled absolute error of 0.074 e^- (scaling factor = 1.42), and a correlation coefficient (r) of 0.970. Merz examined a much larger data set consisting of 1007 atomic charges derived exclusively from HF/6–31G* calculations on amino acids and monosaccharides. He again found much better agreement with MNDO wave functions (r = 0.96) than AM1 (r = 0.81) and an optimum scaling factor of 1.29 for the MNDO charges. Merz also demonstrated that the best scaling factor for MNDO charges is significantly different for amino acids (1.26) than for monosaccharides (1.47). Furthermore, he showed that MNDO, while far superior to AM1, performed extremely poorly for nitrogen (r = 0.50).

One problem with the semiempirical approaches discussed to this point is that they require analytical evaluation of the ESP, a process that easily dominates the CPU time requirements of the entire procedure. There have been a number of attempts to circumvent this CPU bottleneck,[18–22] culminating[23, 24] in a very recent method proposed by Wang and Ford (WF). In the WF method, as in some earlier approaches, the deorthogonalization step is omitted entirely. In addition, the molecular ESP is evaluated via new semiempirical expressions that involve only two-center terms. This approach is approximately 2 orders of magnitude faster than methods that employ semiempirical wave functions and analytic evaluation of the ESP, while achieving reasonable accuracy in the high and low potential regions surrounding the molecule. For a collection of 21 molecules including 156 symmetry unique H, C, N, and O atoms, WF found an average absolute error (relative to HF/6–31G*) of 0.097 e^- and a correlation coefficient of 0.963. When a uniform scaling factor of 1.208 was applied to the calculated charges, the error was reduced to 0.073 e^-. The WF method is currently parametrized for H, C, N, O, F, and Cl, although ESP-derived charges have been reported for only H, C, N, and O.[24]

It seems clear that further progress must be made in the methodology of ESP-derived atomic charge calculations. While *ab initio* methods can in principle achieve any desired level of accuracy, reasonably large basis sets are necessary for acceptable accuracy. Such calculations are not practical for large molecules on a routine basis. Fully semiempirical methods such as the WF approach deal successfully with the CPU time tissue. However, this method is currently parametrized for only a few atoms, uses uncorrelated HF/6−31G* calculations as its reference, and requires a fairly large scaling factor for best agreement with the *ab initio* results, even though it is parametrized against *ab initio* calculations. The use of a scaling factor is always a disadvantage, because one must assume a universal scaling factor for all molecules. Even with a scaling factor, the WF approach still yields an average absolute error of $\sim 0.07e^-$ for a limited set of molecules. In addition, it will clearly be difficult to parametrize for atoms for which *d* orbital participation is important (e.g., hypervalent species and metals).

Ideally one would also like to obtain charges from correlated wave functions, but this would even more severely limit the size of the molecule that could be examined. The importance of electron correlation is apparent in a very recent study by Wampler.[25] Wampler employed a data set consisting of 123 atoms in nine molecules and presented optimum scaling factors for many different computational procedures (HF, MP2, and a variety of density functional approaches) with a large number of basis sets. Taking HF/6−31G* calculations as a reference, he found scaling factors of 0.978 (HF/6−311G**) and 1.086 (MP2/6−311G**). Thus electron correlation, as approximated by the MP2 approach, results in about a 10% change in the ESP-derived charges. This is not a trivial effect, because ESP-derived charges are often quite large (see below).

In this article a new approach for the calculation of ESP-derived charges is presented. It is based on the PRDDO/M method[26−30] and is parametrized against correlated wave functions employing a flexible basis set. Intermediate in computational complexity between a fully *ab initio* and a fully semiempirical approach, the method described here is an order of magnitude faster than the simplest *ab initio* approach (STO-3G) while achieving a degree of accuracy approaching that of the reference calculations against which it is parametrized (*ab initio* MP2/6−31G**). Unlike previous approximate methods, the new method, denoted PESP (parametrized ESP), achieves high accuracy through a direct parametrization of the wave function and analytic (but rapid) calculation of the resultant ESPs. The current PESP program is parametrized for H, C, N, O, F, P, S, Cl, and Br, including hypervalent species for P, S, and Cl. It achieves an overall unscaled average absolute deviation relative to MP2/6−31G** of 0.037 e^- for 820 symmetry unique atoms in 145 molecules, with a correlation coefficient of 0.990.

Elements of PRDDO/M Approach

The method of *partial retention of diatomic differential overlap* (PRDDO) was developed by Halgren and Lipscomb many years ago[26,27] for molecules containing the atoms hydrogen through fluorine and extended by Marynick and Lipscomb[28] through the first transition series. PRDDO is an approximate *ab initio* technique designed to reproduce an *ab initio* minimum basis set calculation in a fraction of the computational time. PRDDO/M[29,30] is a modern version of the method, for which calculations in the 500−1500 orbital range are routine on workstations; calculations approaching 4000 orbitals have been performed[30] on supercomputers. Both PRDDO and PRDDO/M are much closer in spirit and numerical implementation to *ab initio* methods than are typical semiempirical approaches. They are parametrized against *ab initio* calculations (as opposed to experiment) and they retain a significant number of two electron integrals. For instance, virtually all one- and two-center integrals in PRDDO/M are calculated analytically, while three-center integrals are evaluated to an accuracy of about three decimal places.

The PRDDO/M method employs a minimum basis set of Slater orbitals, with two important exceptions: the main-group elements Al−Cl may optionally have a valence 3d orbital, and the atoms Sc-Br employ a contracted double-zeta set of d orbitals (a single Slater 3d function is a notoriously bad representation of a transition metal d orbital).[31] It makes use of three mathematically equivalent basis sets. The first (ρ) consists of pure Slater orbitals. The second (ϕ) is a Schmidt orthogonalized, principle axes basis (see below). The third basis (χ) consists of orthogonal AOs (OAOs), which are obtained from the ϕ basis via a Löwdin

symmetric orthogonalization.[32]

$$\rho(\text{Slater}) \xrightarrow{\mathbf{B}} \phi \xrightarrow{S_\phi^{-1/2}} \chi(\text{orthogonal}). \qquad (4)$$

Here **B** is a block-diagonal matrix that simultaneously Schmidt orthogonalizes the Slater orbitals and rotates the p (or d) components to local principal axes. The **B** transformation is unique in an anisotropic environment and eliminates most of the need for drastic spherical averaging typical of approximate molecular orbital methods.

In the χ basis, integrals of the form $(\chi_i \chi_j|\chi_k \chi_l)$ are vanishingly small when $i \neq j \neq k \neq l$ and may be neglected without significant loss of accuracy.[26] This is the basis for the formal reduction of computational dependence on basis set seize from N^4 to N^3. However, even integrals of the form $(\chi_i \chi_j|\chi_k^2)$ are small in the χ basis, and only a fraction of the integrals of this form survive. In principle, a PRDDO calculation is nothing more than a HF calculation in the χ basis. However, because an exact $\rho \rightarrow \chi$ transformation would require the complete N^4 set of integrals in the ρ basis, approximate transformations must be employed. In some cases, these approximately transformed integrals are then adjusted via parametrizations against *reference parameter set* (RPS) of *ab initio* calculations. PRDDO/M introduces three important modifications of the original PRDDO method: a more sophisticated parametrization is performed, valence electron only calculations using modified frozen core potentials (FCPs) are allowed,[30] and calculations can be performed in a basis that is almost, but not quite, orthogonal (the not quite OAO or NQOAO basis).[29] The NQOAO option makes very large calculations possible with essentially no loss in accuracy, especially with regard to the charge distribution.

The PRDDO/M method employs four parameters for each atom (five for the atoms Al–Cl with a valence d orbital). These parameters modify integrals in the χ basis as follows:

$$\left(\chi_{1sH}\chi_{jA}|\chi_{jA}^2\right) \Rightarrow p_1\left(\chi_{1sH}\chi_{jA}|\chi_{jA}^2\right), \qquad (5)$$

$$\left(\chi_{iB}\chi_{jA}|\chi_{jA}^2\right) \Rightarrow p_2 c_1\left(\chi_{iB}\chi_{jA}|\chi_{jA}^2\right)$$

$$c_1 = \left(\chi_{jA}^2|\chi_{jA}^2\right)/\left(\phi_{jA}^2|\phi_{jA}^2\right), \qquad (6)$$

$$(\chi_{iA}^2|\chi_{iA}^2) \Rightarrow p_3(\chi_{iA}^2|\chi_{iA}^2) + (1 - p_3)(\phi_{iA}^2|\phi_{iA}^2), \qquad (7)$$

$$\left(\chi_{iB}\chi_{1sH_A}|\chi_{1sH_A}^2\right) \Rightarrow p_4 c_2\left(\chi_{iB}\chi_{1sH_A}|\chi_{1sH_A}^2\right),$$

$$c_2 = \left(\chi_{1sH_A}^2|\chi_{1sH_A}^2\right)/\left(\phi_{1sH_A}^2|\phi_{1sH_A}^2\right), \qquad (8)$$

$$\left(\chi_{3d_iB}\chi_{1sH_A}|\chi_{1sH_A}^2\right) \Rightarrow p_5 c_2\left(\chi_{3dB}\chi_{1sH_A}|\chi_{1sH_A}^2\right). \qquad (9)$$

For eqs. (5), (6), and (8) i and j run over the valence s and p orbitals, while i runs over the 3p orbitals in eq. (7) and the 3d orbitals in eq. (9). All parameters are either of a one-center nature or involve two-center integrals that fall off rapidly with distance (i.e., two-center Coulomb integrals are *not* parametrized). In the most recent[29–31] versions of the method, the RPS was defined by *ab initio* STO-3G calculations and the error function minimized included internuclear distances, Mulliken charges, and a limited set of energetics. Further details of the method may be found in the original articles.[29–31] It is important to note, however, that the PRDDO/M parametrization is extremely mild: unparametrized PRDDO/M calculations yield reasonable charge distributions and energetics compared to the reference *ab initio* STO-3G calculations and are from 20 to 50 times faster. Indeed, the current implementation of PRDDO and PRDDO/M contains no parameters for the transition metals, Sc-Zn, yet both methods have proven very useful for the study of a wide variety of problems in transition metal chemistry.[33]

PESP Method

MODIFICATIONS OF PRDDO/M METHOD

The essence of the PESP modifications to the PRDDO/M method is extremely simple. A PESP calculation is simply a PRDDO/M/FCP calculation with a new set of parameters (Table I) chosen to minimize an error function based upon ESP-derived atomic charges:

$$\varepsilon = \sum_i |q_i^{ab\ initio} - q_i^{\text{PRDDO}}|. \qquad (10)$$

All *ab initio* charges were obtained from MP2/6–31G** calculations using the program Gaussian 94[34] on an NEX SX-3 computer. All geometries were identical for the *ab initio* and PESP calculations and were obtained from either an experiment[35] or optimizations at a well-defined theoretical level. Geometry optimizations were performed on an IBM RS-6000 model 250 computer using the program GAMESS.[36] Most optimizations were performed at the 6–31G* (6d) level; however, a vari-

TABLE I.
Final Parameters.

Atom	Exponents 1s	2sp	3sp	3d	4sp	$p1$	$p3$
H	1.322						
C		1.511				0.465	1.675
N		1.941				0.831	0.757
O		2.115				0.578	1.552
F		2.397					1.066
P	8.700	5.091	1.779	1.905			1.080
S	14.76	4.611	2.110	2.583			0.921
Cl	13.33	5.227	2.342	2.775			0.006
Br	7.479	10.03	8.286	6.516	2.544		3.084

Blank entries indicate that the parameter is unchanged from the original PRDDO/M method.

ety of other levels, including STO-3G, MNDO, and 3−21G were included in an effort to make the parametrization largely independent of the theoretical model used to obtain the geometry. Because ESP-derived charges are not in principle rotationally invariant, the exact same molecular orientation was employed in the *ab initio* and the PESP calculations.

The general approach to parametrization was as follows. Initially, the parameters for hydrogen, carbon, oxygen, and nitrogen were optimized simultaneously for a group of 45 molecules defined in Table II. Only the charges of symmetry-unique atoms were considered. The parameters for these atoms were then fixed. For each additional atom, a RPS of *ab initio* charges was defined (Tables

TABLE II.
Molecules in C, H, N, and O RPS and Statistical Results.

Molecule	Geometry	Molecule	Geometry
Alanine dimer	6−31G*	Alanine-glycine	3−21G
Adenine	MNDO	Alanine	6−31G*
2-Aminopropanal	6−31G*	Allose	6−31G*
Aminobenzene	6−31G*	Asparagine	6−31G*
Acetylene	Experiment	Ethylene	Experiment
2(3H)Furanone	6−31G*	$H_2C=NH$	Experiment
Formaldehyde	Experiment	Methylamine	6−31G*
Nitromethane	6−31G*	Methanol	6−31G*
Methane	Experiment	Formamide	6−31G*
Carbon monoxide	Experiment	Carbon dioxide	Experiment
Ethylene Oxide	6−31G*	Formic acid	6−31G*
Furan	6−31G*	Glutamine	3−21G
Guanine	MNDO	Carbonic acid	6−31G*
Water	Experiment	Aminoacetylene	6−31G*
Hydroxyacetylene	6−31G*	Hydrogen cyanide	Experiment
HN=NOH	6−31G*	HN=O	Experiment
Nitric acid	6−31G*	H_2NOH	6−31G*
H_2NOOH	6−31G*	Ammonia	Experiment
Phenol	STO-3G	Phenylalanine	STO-3G
Pyridine	6−31G*	Pyridine *n*-oxide	6−31G*
Pyrizole	6−31G*	Ribose	6−31G*
Tetrahydrofuran	6−31G*	Thymine	3−21G
Uracil	6−31G*		

Atom	No. Unique Atoms	Average Abs. Charge	Abs. Error	Signed Error
H	171	0.219	0.020	−0.002
C	111	0.356	0.064	−0.006
N	43	0.712	0.056	0.017
O	61	0.500	0.030	0.002

Overall average deviation = 0.038 and overall correlation coefficient (r) = 0.991.

TABLE III.
Molecules in Fluorine RPS and Statistical Results.

Molecule	Geometry	Molecule	Geometry
1,1-Difluoroethylene	6–31G*	1,2-Difluoroethylene	6–31G*
$O=CF_2$	STO-3G	Carbon tetrafluoride	Experiment
Fluoromethane	Experiment	Difluoroacetylene	6–31G*
Trifluoroacetic acid	6–31G*	Fluorobenzene	6–31G*
Fluorocyanide	6–31G*	2-Fluorofuran	MNDO
Parafluoropyridine	STO-3G	H(CO)F	Experiment
Hydrogen fluoride	Experiment	Nitrogen trifluoride	Experiment
O=NF	Experiment	Oxygen difluoride	Experiment

Atom	No. Unique Atoms	Average Abs. Charge	Average Abs. Error	Average Signed Error
H	15	0.183	0.013	0.008
C	21	0.367	0.052	−0.005
N	4	0.331	0.029	0.012
O	7	0.300	0.026	−0.004
F	17	0.155	0.020	−0.004

Overall average deviation = 0.030 and overall correlation coefficient (r) = 0.989.

III–VII) and the error function [eq. (10)] was minimized. Atoms were then parametrized sequentially in the order fluorine, chlorine, phosphorus, sulfur, and bromine. Again, once the parameters for a given atom were optimized, they were not allowed to vary when a new atom was added. A simplex procedure was used for optimization of the parameters.

TABLE IV.
Molecules in Chlorine RPS and Statistical Results.

Molecule	Geometry	Molecule	Geometry
1,1-Dichloroethylene	6–31G*	1,1-Chlorofluoethylene	6–31G*
1,2-Dichloroethylene	6–31G*	Trichloroacetic acid	6–31G*
Carbon tetrachloride	Experiment	Chloromethane	Experiment
Chorobenzene	6–31G*	Chloroacetylene	6–31G*
Chlorocyanide	6–31G*	Chlorine fluoride	Experiment
Chlorine trifluoride	Experiment	Chlorine pentafluoride	Experiment
2-Chlorofuran	MNDO	Chlorate anion	6–31G*
Perchlorate anion	6–31G*	Parachloropyridine	STO-3G
Fluorochloromethane	6–31G*	Chloric acid	6–31G*
Perchloric acid	6–31G*	HOCl	Experiment
Nitrogen trichloride	Experiment	Oxygen dichloride	Experiment
Phosgene	Experiment		

Atom	No. Unique Atoms	Average Abs. Charge	Average Abs. Error	Average Signed Error
H	19	0.207	0.013	0.009
C	25	0.245	0.040	0.005
N	3	0.415	0.036	−0.017
O	13	0.367	0.035	−0.013
F	7	0.130	0.018	−0.004
Cl	24	0.240	0.042	0.000

Overall average deviation = 0.033 and overall correlation coefficient (r) = 0.990.

TABLE V.
Molecules in Phosphorus RPS and Statistical Results.

Molecule	Geometry	Molecule	Geometry
$CH_3N=PF_3$	6–31G*	$CH_3P=NH$	6–31G*
Cyanophosphine	MNDO	Aminophosphine	6–31G*
Phosphoric acid	6–31G*	$HC \equiv CPH_2$	6–31G*
$P(NH_2)(H)(F)$	6–31G*	2-phosphirene	6–31G*
Dimethylphosphate anion	6–31G*	$O=PCl_3$	6–31G*
$O=PF_3$	6–31G*	Phosphorus trichloride	Experiment
Phosphorus trifluoride	Experiment	Phosphorus pentachloride	MNDO
Phosphorus pentafluoride	Experiment	Phosphole	6–31G*
$P(CH_3)_2(OCH_3)$	6–31G*	$P(CH_3)(F)(OH)$	6–31G*
$P(OH)_3$	6–31G*	Vinylphosphine	STO-3G

Atom	No. Unique Atoms	Average Abs. Charge	Average Abs. Error	Average Signed Error
H	35	0.173	0.022	0.013
C	14	0.287	0.065	−0.043
N	5	0.616	0.035	−0.002
O	9	0.481	0.053	0.053
F	7	0.168	0.034	−0.033
P	20	0.439	0.066	−0.008
Cl	4	0.080	0.034	0.031

Overall average deviation = 0.043 and overall correlation coefficient (r) = 0.989.

TABLE VI.
Molecules in Sulfur RPS and Statistical Results.

Molecule	Geometry	Molecule	Geometry
$CH_3C(O)SH$	6–31G*	$CH_3C(S)OH$	6–31G*
Methylmethoxysulfide	6–31G*	Dimethylsulfide	6–31G*
Fluoromethylsulfide	6–31G*	Chloroaminosulfide	MNDO
CS	Experiment	Hydrogen sulfide	Experiment
Sulfuric acid	6–31G*	Vinylhydrogensulfide	MNDO
Cyanohydrogensulfide	6–31G*	Aminohydrogensulfide	MNDO
Bisulfate anion	6–31G*	$O=SCl_4$	6–31G*
OSF_4	6–31G*	Phenylhydrogensulfide	6–31G*
Thiirene	6–31G*	$ClSCH_3$	6–31G*
Sulfur dioxide	Experiment	Sulfur trioxide	Experiment
Thiophene	6–31G*		

Atom	No. Unique Atoms	Average Abs. Charge	Average Abs. Error	Average Signed Error
H	37	0.184	0.019	0.013
C	32	0.252	0.058	−0.021
N	3	0.564	0.031	0.031
O	13	0.452	0.019	0.001
F	2	0.127	0.084	−0.084
S	21	0.392	0.070	−0.009
Cl	3	0.082	0.034	0.022

Overall average deviation = 0.041 and overall correlation coefficient (r) = 0.984.

TABLE VII.
Molecules in Bromine RPS and Statistical Results.

Molecule	Geometry	Molecule	Geometry
1,1-Dibromoethylene	6–31G*	1,1-Fluorobromoethylene	6–31G*
1,2-Dibromoethylene	6–31G*	Bromobenzene	6–31G*
Bromoacetylene	6–31G*	Bromine chloride	6–31G*
Bromine cyanide	6–31G*	Bromine fluoride	6–31G*
2-Bromofuran	MNDO	Parabromopyridine	STO-3G
Brominehydrogen sulfide	6–31G*	Tribromoacetic acid	6–31G*
Tribromomethane	6–31G*	Bromochloromethane	3–21G
Bromomethane	6–31G*	H(CO)Br	6–31G*
Aminobromine	6–31G*	NOBr	6–31G*
Oxygen dibromide	6–31G*	Bromophosphine	6–31G*

Atom	No. Unique Atoms	Average Abs. Charge	Average Abs. Error	Average Signed Error
H	22	0.195	0.016	0.011
C	25	0.322	0.044	0.009
N	4	0.445	0.042	−0.023
O	6	0.279	0.047	−0.044
F	2	0.113	0.042	−0.042
P	1	0.080	0.084	−0.084
S	1	0.123	0.020	0.020
Cl	2	0.034	0.044	0.040
Br	21	0.098	0.030	−0.005

Overall average deviation = 0.034 and overall correlation coefficient (r) = 0.989.

EVALUATION OF ESP

For modern *ab initio* codes, the evaluation of the ESP at the HF level is a trivial computational task compared to the calculation of the wave function. (However, at the MP2 level the ESP calculation can be time consuming.) For semiempirical methods that explicitly deorthogonalize the wave function and calculate the ESPs analytically (usually via a Gaussian expansion) the opposite is true: the ESP part of the calculation dominates the total computational time. In the current approach, an effort was made to develop a balanced algorithm in which the ESP part of the calculation requires roughly the same amount of computational effort as the wave function calculation. This is not a trivial task, because the ESP must be evaluated over Slater orbitals (or equivalently over a high quality Gaussian expansion such as STO-6G). Such an algorithm is described here.

The nuclear part of the ESP is calculated exactly, and all two-center integrals are evaluated analytically over the Slater orbital basis set. Three-center terms, which are not analytic over Slater orbitals, are approximated as follows. First, the integrals are prescreened. For the ESP at the point C, the required integral in the point charge approximation is

$$\left(\rho_{iA}\,\rho_{jB}\left|\frac{1}{r_{1C}}\right.\right) \approx S_{ij}/r, \tag{11}$$

where S_{ij} is the Slater orbital overlap integral and r is the distance from point C to the centroid of the ij charge distribution. Taking $r = 1$ provides an upper bound to the point charge approximation for this integral. ESPs at points closer than 1 bohr from a nucleus are of no practical importance for ESP-derived charges or most other properties related to ESPs. Thus, the approximate upper bound for the contribution (E^A_{ijC}) of this integral to the ESP is

$$E^A_{ijC} = S_{ij}P_{ij}. \tag{12}$$

When E^A_{ijC} is less than 1.0×10^{-9} for the s–s integral of a given nsp–n′sp shell pair, these integrals are neglected for *all* points C. For integrals involving d orbitals, only the σ component of the overlap integral in a local diatomic coordinate system is checked. All three-center Slater orbital integrals that survive this test are evaluated by var-

iable length Gaussian expansions of the Slater orbitals, corrected by the ratio of the Slater overlap integral to the corresponding Gaussian overlap.

$$\left(\rho_{iA}\,\rho_{jB}\left|\frac{1}{r_{1C}}\right.\right) \approx \left(g_i^{(N)}g_j^{(N)}\left|\frac{1}{r_{1C}}\right.\right)\frac{S_{i,j}^{\rho}}{S_{i,j}^{g(N)}}. \quad (13)$$

Here $S_{i,j}^{\rho}$ is the Slater orbital overlap integral, $S_{i,j}^{g(N)}$ is the same overlap integral calculated via an N Gaussian expansion, and $g_i^{(N)}$ represents an N Gaussian expansion of a Slater orbital. The length of the Gaussian expansion is determined by an estimate of the point charge (PC) contribution to the ESP, E_{ijC}^{PC}, defined as:

$$P_{ij}\left(\rho_{iA}\,\rho_{jB}\left|\frac{1}{r_{1C}}\right.\right) \approx E_{ijC}^{\mathrm{PC}} = P_{ij}S_{ij}/r_{\min},$$

$$r_{\min} = \min(R_{AC}, R_{BC}). \quad (14)$$

When $E_{ijC}^{\mathrm{PC}} < 10^{-4}$, $N = 2$. For $10^{-3} > E_{ijC}^{\mathrm{PC}} \leq 10^{-4}$, $N = 3$. Otherwise, $N = 4$. This procedure works because the major error introduced when substituting short Gaussian expansions for Slater orbitals is in the orbital overlap, which is underestimated by the Gaussian expansion. The ratio of the overlap integrals in eq. (13) corrects this deficiency nearly quantitatively and results in ESPs that agree very well with STO-6G values, even though the vast majority of three-center terms are calculated at the $N = 2$ level. For the entire data set of 820 atoms described here, the ESP-derived charges calculated with the above approach exhibit an average absolute deviation of $0.0016e^-$ relative to those calculated with a full STO-6G expansion. In terms of the actual electrostatic potentials, the average absolute error for this procedure relative to a full STO-6G expansion is 0.054 kcal/mol. This error was based on a total of 84,795 points required to sample the ESP for all 145 molecules in the data set.

Statistical Results for ESP-Derived Charges

HYDROGEN, CARBON, NITROGEN, AND OXYGEN

Only the 1s exponent was taken as a free parameter for hydrogen. For carbon, nitrogen, and oxygen, the valence 2sp exponents were optimized. In addition, parameters p_1 and p_3 (defined above) were included in the parametrization. This results in a total of 10 parameters to be defined by the 386 atom data set. The CHNO RPS (Table II) consists of four amino acids, two dipeptides, two sugars, and four DNA/RNA bases. In addition, a wide variety of inorganic and organic species representative of many different bonding situations were included. The correlation coefficient (here and in all subsequent statistics the *ab initio* charges are taken as the dependent variable) is 0.991. The maximum error for all 386 observations is -0.269 e^- for one carbon in the alanine-glycine dipeptide. The average absolute error is 0.038 e^-. As demonstrated by the very small average signed errors (Table II), the residual errors are essentially random. Linear regression of the form,

$$q_{\mathrm{MP2}} = mq_{\mathrm{PESP}} + b, \quad (15)$$

yields $m = 0.992$ and $b = 0.0002$. Regression with a uniform scaling factor,

$$q_{\mathrm{MP2}} = mq_{\mathrm{PESP}}, \quad (16)$$

yields the same value for m. The fact that m is essentially unity for eq. (16) implies that uniform scaling of the calculated charges is unnecessary. This is an important and general result: unlike semiempirical methods, PESP charges do not require a uniform scaling factor to improve agreement with the underlying *ab initio* results.

FLUORINE

The fluorine RPS (Table III) consists of 16 molecules with 64 symmetry unique atoms and 17 fluorines. Only the p_3 and the 2sp exponent were optimized, with an overall average absolute deviation of 0.030 e^- for the entire data set and 0.020 e^- for fluorine. Liner regression yields $m = 0.976$ and $b = 0.0008$. The correlation coefficient is 0.989. It is especially encouraging that the average absolute deviations for hydrogen, carbon, nitrogen, and oxygen in the fluorine RPS are actually somewhat *smaller* than those found in the CHNO RPS. This is remarkable, given that the CHNO parameters were not varied during the fluorine parametrization but were previously determined from the CHNO RPS. Similar behavior is seen in many of the other data sets (see below) and suggests that the CHNO parametrization is quite general. The largest error for any atom in the fluorine RPS is 0.214 e^- (carbon in CF_4), while the largest error for fluorine is -0.053 e^- (fluorine in CF_4).

CHLORINE, PHOSPHORUS, AND SULFUR

For chlorine, phosphorus, and sulfur the parameters optimized included p_3 and the 1s, 2sp, 3sp, and 3d exponents. All of these atoms have a set of 3d orbitals explicitly included in their basis set. It is important to note that variation of the inner shell elements for all of the remaining atoms only affects the calculation of the modified FCP. The valence electron only implementation of PRDDO/M uses the electron repulsion integrals involving core orbitals to construct the FCP, then discards the core orbitals. Therefore, using core exponents as optimization parameters is equivalent to modifying the FCP.

The chlorine RPS (Table IV) consists of 23 molecules with 91 total atoms and 24 chlorines. It includes six hypervalent species. The overall average absolute deviation is 0.033 e^- and the average absolute deviation for chlorine is 0.042 e^-. The linear regression parameters are $m = 0.947$ and $b = -0.0003$, with a correlation coefficient of 0.990. The largest error for any atom in this data set is 0.247 e^- for the chlorine in $HClO_4$. The errors for the nonchlorine atoms are again slightly smaller than or comparable to those found in the previous data sets.

The phosphorus RPS (Table V) consists of 20 molecules, seven of which are hypervalent species. There are 87 total atoms and 20 phosphorus atoms. The overall average absolute deviation for all atoms is 0.043 e^-, while that of phosphorus is 0.066 e^-. The linear regression parameters are $m = 1.020$ and $b = -0.0006$, and the correlation coefficient is 0.989. The largest error for any atom in this data set is -0.214 e^- for carbon in the dimethylphosphate anion, while the largest error for phosphorus is 0.157 e^- in PF_3. All of the nonphosphorus atoms exhibit similar to those seen in the previous tables. While the overall error for phosphorus is somewhat larger than that found for the other atoms, phosphorus exhibits a very wide range of charges at the *ab initio* level (see below). Therefore, the relative errors are quite small.

The sulfur RPS (Table VI) has 21 molecules with six hypervalent species. There are 101 total atoms and 21 sulfur atoms. The overall average absolute deviation is 0.041, while that of sulfur is 0.070. The linear regression parameters are $m = 0.975$ and $b = 0.0016$. The correlation coefficient is 0.984. The largest error in this data set, and the largest error for all of the molecules in this article, is 0.356 e^- for the sulfur in OSF_4.

BROMINE

For bromine, the 1s, 2sp, 3sp, 3d, and 4sp exponents were varied; for consistency with the *ab initio* reference calculations the 3d orbital was taken as an STO-6G expansion of a single Slater 3d (as opposed to the contracted double-zeta representation, which is the default PRDDO/M description). The parameter p_3 was also included in the optimization. It is likely that fewer parameters are actually needed to achieve satisfactory agreement with the *ab initio* RPS; however, it became clear that variation of the 1s exponent was absolutely necessary and it is difficult to justify varying the 1s exponent while holding other inner shell exponents fixed.

The bromine RPS consists of 20 molecules with 84 atoms and 21 bromines. No hypervalent species were included in this RPS, because the current implementation of PRDDO/M does not allow valence 4d orbitals for bromine. The overall average absolute deviation is 0.034 e^-, while that of bromine if 0.030 e^-. The linear regression parameters are $m = 0.944$ and $b = -0.0006$. The correlation coefficient is 0.989. The largest error in this data set is 0.200e^- for one carbon in CBr_3COOH. The largest bromine error is 0.099 e^- in NOBr.

OVERALL CHARGE STATISTICS

The overall charge statistics for all 820 atoms in 145 molecules are presented in Table VIII. An overall absolute deviation of 0.037 e^- was achieved, with a correlation coefficient of 0.990. Linear regression on the entire data set yields $m = 0.986$ and $b = 0.0003$. A plot of q_{MP2} versus q_{PESP} is presented in Figure 1. Table VIII also presents the range of ESP-derived charges (from the MP2 calculations) for each atom, and it can be seen that these charges vary widely depending on the molecular environment. Finally, this table presents correlation coefficients for each atom taken separately. For instance, linear regression of 299 hydrogen charges with the MP2 values as the dependent variable yields $r = 0.987$. The least significant correlation occurs for fluorine, where $r = 0.920$; however, fluorine has the *smallest* average absolute deviation of any of the nonhydrogen atoms. This apparent inconsistency is easily resolved when the range of fluorine charges is considered. Fluroine exhibits the narrowest range of charges for any nonhydrogen atom, and this al-

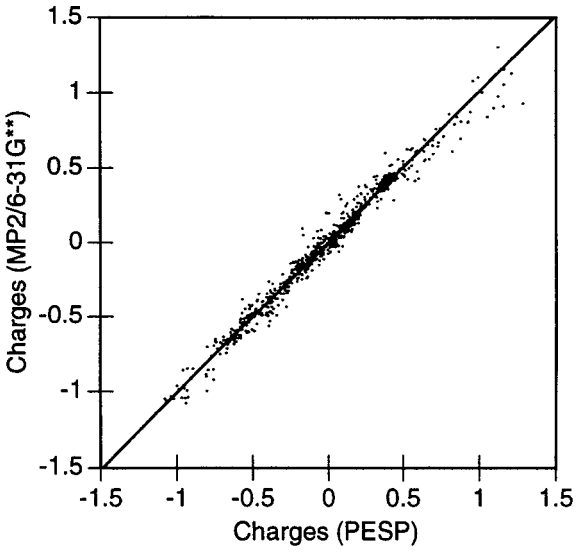

FIGURE 1. Plot of PESP atomic charges vs. ESP-derived atomic charges from *ab initio* MP2/6–31G** wave functions for all 820 atoms in the data set.

lows for high absolute accuracy with a relatively low correlation coefficient.

Earlier it was noted that the parameters derived for hydrogen, carbon, nitrogen, and oxygen clearly yielded excellent statistical results for molecules not in the CHNO RPS. This is demonstrated in Table IX, where the statistics for these four atoms derived from all of the molecules *not* in the CHNO RPS are presented. This data set consists of 100 molecules, each with at least one atom than C, H, N, and O. There are a total of 302 unique C, H, N, and O atoms in this data set. The overall average absolute deviation, average absolute deviations for each atom, and correlation coefficient are very similar to those found for the molecules used to derive the parameters for these atoms (Table II). Indeed, in two cases (carbon and nitrogen) the statistical agreement is actually somewhat better. Because these molecules were used in parameterizations of atoms other than C, H, N, and O, this is

TABLE VIII.
Overall Charge Statistics for All Molecules.

Atom	No. Unique Atoms	Average Abs. Charge	Average Abs. Error	Average Signed Error	Range	*r*
H	299	0.205	0.019	0.003	−0.079 / 0.466	0.987
C	218	0.325	0.062	−0.005	−1.044 / 1.305	0.979
N	62	0.641	0.048	0.013	−1.078 / 0.842	0.989
O	109	0.452	0.033	0.002	−0.765 / 0.122	0.970
F	35	0.148	0.026	−0.015	−0.432 / − 0.007	0.920
P	21	0.422	0.064	−0.009	−0.859 / 0.810	0.990
S	22	0.380	0.069	−0.006	−0.341 / 1.305	0.979
Cl	33	0.194	0.042	0.008	−0.145 / 0.978	0.981
Br	21	0.098	0.030	−0.005	−0.229 / 0.222	0.946

Overall average deviation = 0.037 and overall correlation coefficient (r) = 0.990.

TABLE IX.
Statistics for Carbon, Hydrogen, Nitrogen, and Oxygen for Molecules Not in CHNO RPS.

Atom	No. Unique Atoms	Average Abs. Charge	Average Unsigned Error	Average Signed Error
H	128	0.186	0.018	0.011
C	107	0.294	0.050	−0.008
N	19	0.480	0.035	0.001
O	48	0.391	0.034	0.001

Overall average deviation = 0.036 and overall correlation coefficient (r) = 0.988.

not a truly independent test of the method. However, these results, coupled with the large size and diversity of the RPS, clearly demonstrate the usefulness of the PESP procedure.

Dipole Moments

Another important measure of the accuracy of ESP-derived charges is the dipole moment derived from them. Table X lists the dipole moments calculated three ways for all molecules (excluding charged species and those with $\mu = 0$ by symmetry.) Table XI summarizes the statistical analysis for the data in Table X. Dipole moments were calculated directly from the MP2 wave function, from the MP2 ESP-derived atomic charges (denoted MP2-AC), and from the PESP atomic charges (denoted PESP-AC). It was not possible to calculate dipole moments directly from the PRDDO/M/FCP wave function, because this capability is not yet programmed into the method. As seen in Table XI, at the MP2 level dipole moments calculated from the ESP-derived charges correlate essentially perfectly with those calculated directly from the wave function. The PESP dipole moments correlate with the MP2-AC dipole moments reasonably well, with a correlation coefficient of 0.952 and an average absolute deviation of 0.28 D. Application of a uniform scaling factor of 0.92 reduces this error slightly (to 0.24 D), but scaling is not recommended because the residual errors in the charges are nearly random.

CPU Time Comparisons

Table XII lists CPU time requirements for PESP calculations and various levels of *ab initio* theory using Gaussian 94 on an HP 735 workstation. For PESP calculations on small molecules, the ESP part of the calculation dominates ($\sim 80\%$) the CPU time. For larger molecules the wave function calculation begins to dominate. For instance, 70% of the time required to calculate the PESP charges for the DNA dimer GA (including the complementary base pairs, sugars, and phosphates) is in the wave function calculation. The wave function calculation dominates the CPU time of *ab initio* ESP calculations at any level of the basis set. For large molecules, PESP calculations are an order of magnitude faster than *ab initio* calculations at the STO-3G level and are more than 2 orders of magnitude

TABLE X.
Calculated Dipole Moments for All Molecules.

Molecule	MP2[a]	MP2-AC[b]	PRDDO-AC[c]
1,1-Chlorofluoroethylene	1.27	1.31	1.63
1,1-Dibromoethylene	1.33	1.38	1.81
1,1-Dichloroethylene	1.45	1.51	1.68
1,1-Difluoroethylene	0.97	1.00	1.50
1,1-Fluorobromoethylene	1.23	1.27	1.75
1,2-Dibromoethylene	1.82	2.04	2.73
1,2-Dichloroethylene	2.03	2.15	2.67
1,2-Difluoroethylene	2.21	2.16	2.31
2-Bromofuran	1.02	1.08	1.53
2-Chlorofuran	1.20	1.24	1.56
2-Fluorofuran	0.89	0.89	1.18
$C_4H_4O_2$	3.68	3.71	4.01
$CH_3C(O)SH$	1.33	1.36	1.70
$CH_3C(S)OH$	1.90	1.99	1.53
$CH_3N=PF_3$	1.51	1.54	2.78
$CH_3P=NH$	0.88	0.94	1.12
CS	2.03	2.00	1.88
H(CO)Br	1.74	1.84	2.02
H(CO)F	1.95	1.96	1.97
$H_2C=NH$	2.00	1.98	2.03
H_2NOH	3.31	3.28	3.49
H_2NOOH	1.44	1.38	1.70
$HC \equiv CPH_2$	0.33	0.36	0.69
HN=O	1.59	1.54	1.87
HN=OH	1.95	1.91	2.08
HOCl	1.75	1.81	1.93
2-Phosphirene	0.64	1.59	1.50
NOBr	2.52	2.62	1.39
NOF	1.54	1.60	1.11
$O=CF_2$	1.04	1.01	0.81
$O=PCl_3$	2.05	2.03	1.68
$O=PF_3$	1.77	1.78	0.72
$O=SCl_4$	1.16	1.09	1.60
$O=SF_4$	1.26	1.23	0.35
$P(CH_3)(F)(OH)$	1.42	1.41	1.96
$P(CH_3)_2(OCH_3)$	0.10	0.04	0.42
$P(NH_2)(H)(F)$	1.97	1.95	2.36
$P(OH)_3$	3.36	3.24	3.29
Thiirene	2.38	2.48	2.36
Adenine	2.46	2.46	2.30
Alanine	1.94	1.94	2.29
2-Aminopropanal	2.65	2.66	2.99
Alanine dimer	2.25	2.22	2.17
Alanine-glycine	3.01	3.00	2.81
Allose	4.06	4.06	4.58
Aminoacetylene	1.73	1.80	1.79
Aminobenzene	1.58	1.66	1.68
Aminobromine	1.96	2.12	2.33
Aminohydrogensulfide	2.66	2.66	2.82
Aminophosphine	0.90	0.88	0.81
Ammonia	1.81	1.86	1.91
Asparagine	4.66	4.65	4.96

(Continued)

TABLE X.
(Continued)

Molecule	MP2[a]	MP2-AC[b]	PRDDO-AC[c]
Bromoacetylene	0.46	0.51	0.72
Bromine chloride	0.42	0.30	0.56[4]
Bromine cyanide	2.92	2.77	2.57
Bromine fluoride	1.26	1.10	1.68
Brominehydrogensulfide	1.27	1.43	1.91
Bromobenzene	1.86	1.97	2.42
Bromochloromethane	1.81	1.99	2.22
Bromomethane	1.96	2.03	2.42
Bromophosphine	1.81	1.91	2.17
Carbon monoxide	0.24	0.18	0.45
Carbonic acid	0.04	0.09	0.15[d]
Chloric acid	3.84	3.89	4.30
Chlorine fluroride	1.01	0.93	1.37
Chlorine pentafluoride	0.71	0.73	0.64
Chlorine trifluoride	0.76	0.74	0.53
Chloroaminosulfide	2.48	2.52	2.47
Chlorocyanide	2.59	2.47	2.21
Chloromethane	1.98	2.04	2.33
$ClSCH_3$	2.21	2.28	2.89
Chloroacetylene	0.64	0.67	0.92
Chlorobenzene	1.85	1.91	2.31
Cyanohydrogensulfide	3.22	3.14	3.18
Cyanophosphine	3.85	3.85	3.83
Dimethylsulfide	1.79	1.82	2.20
Ethylene oxide	1.93	1.95	1.70
Fluorine cyanide	2.26	2.24	1.89
Fluorobenzene	1.31	1.34	1.76
Fluorochloromethane	1.89	1.96	2.02
Fluoromethane	1.79	1.81	1.86
Fluoromethylsulfide	2.10	2.12	3.22
Formaldehyde	2.17	2.17	2.19
Formamiade	3.67	3.65	3.73
Formic acid	1.36	1.35	1.27
Furan	0.58	0.57	0.81
Glutamine	3.24	3.26	3.49
Guanine	6.02	6.04	6.44
Hydroxyacetylene	1.77	1.83	1.93
Hydrogen cyanide	2.91	2.86	2.64
Hydrogen fluoride	1.88	1.90	1.87
Hydrogen sulfide	1.37	1.52	1.44
Methanol	1.71	1.70	1.77
Methylamine	1.43	1.41	1.60
Methylmethoxysulfide	3.89	3.90	4.36
Nitric acid	2.50	2.45	3.02
Nitrogen trichloride	0.71	0.69	0.78
Nitrogen trifluoride	0.05	0.06	0.06
Nitromethane	3.27	3.27	4.00
Oxygen dibromide	0.98	0.94	1.16
Oxygen dichloride	0.85	0.74	0.74
Oxygen difluoride	0.23	0.23	0.22

(Continued)

TABLE X.
(Continued)

Molecule	MP2[a]	MP2-AC[b]	PRDDO-AC[c]
Parabromopyridine	0.81	0.71	0.45
Parachloropyridine	0.58	0.53	0.43
Parafluoropyridine	1.00	0.98	0.83
Perchloric acid	2.53	2.52	2.94
Phenol	1.41	1.42	1.54
Phenylalanine	2.37	2.40	2.51
Phenylhydrogensulfide	1.33	1.37	1.54
Phosgene	0.91	0.81	1.29
Phosphole	0.96	0.94	1.13
Phosphoric acid	0.68	0.67	1.03
Phosphorus trichloride	0.87	0.93	0.47
Phosphorus trifluoride	1.02	1.05	1.63
Pyridine	2.23	2.23	2.46
Pyridine *n*-oxide	4.03	4.06	4.95
Pyrizole	2.28	2.27	2.38
Ribose	2.85	2.86	2.95
Sulfur dioxide	1.64	1.82	1.75
Sulfuric acid	3.99	3.93	4.00
Tetrahydrofuran	1.79	1.78	1.82
Thiophene	0.57	0.53	1.26
Thymine	3.95	3.99	4.09
Tribromoacetic acid	1.85	1.86	1.91
Tribromomethane	0.95	1.14	1.50
Trichloroacetic acid	2.12	2.13	2.23
Trifluoroacetic acid	2.19	2.19	2.34
Uracil	3.95	3.98	4.09
Vinylhydrogensulfide	0.97	0.98	1.23
Vinylphosphine	1.30	1.21	1.24
Water	2.10	2.16	2.16

All values are in debyes. Excludes molecules with $\mu = 0$ by symmetry and charged species.
[a] Calculated from the wave function, 6–31G** basis set.
[b] Calculated from electrostatic potential derived atomic charges.
[c] Dipole moment has incorrect sign.

TABLE XI.
Overall Dipole Moment Statistics.

Methods Correlated (Indep./ Dep.)	Average Abs. Error (D)	Average Scaled Abs. Error[a]	r
MP2-AC / MP2	0.05	0.05 (1.00)	0.998
PESP-AC / MP2	0.31	0.26 (0.91)	0.942
PESP-AC / MP2-AC	0.28	0.24 (0.92)	0.952

Data set consists of 131 molecules containing all parametrized atoms and excluding charged species and molecules with $\mu = 0$ by symmetry. Average absolute dipole moment $= 1.87D$ (calculated from the MP2 wave function).
[a] Best uniform scale factor in parentheses.

TABLE XII.
Comparison of CPU Time Requirements for Wave Function Plus ESP-Derived Charges.

Molecule	Formula	CPU Time (s) PESP	STO-3G	6–31G*	6–31G**	MP2[a]
Alanine	$C_3H_7NO_2$	4	14[b]	480	721	3290
Alanine-Glycine	$C_5H_{10}N_2O_3$	10	37[b]	1560	2370	—
DNA $\begin{pmatrix} G \\ C \end{pmatrix}$	$C_{19}H_{27}N_8O_{11}P$	173[c]	1780[d]	39768	—	—
DNA $\begin{pmatrix} GA \\ CT \end{pmatrix}$	$C_{39}H_{51}N_{15}O_{20}P_2$	1120[c]	10805[d]	—	—	—

Work perfomred on an HP 735 workstation. *Ab initio* calculations were with Gaussian 94 direct SCF.
[a] With the 6–31G** basis set.
[b] For these small molecules, Gaussian 94 keeps the integrals in memory. This results in artificially fast times.
[c] Includes *d* orbitals on phosphorus.
[d] Does not include *d* orbitals on phosphorus.

faster than HF/6–3G*, the lowest level of *ab initio* theory that is commonly considered to be adequate for the determination of ESP-derived charges.

Conclusions

The PESP modifications to the PRDDO/M molecular orbital method achieves a level of accuracy for ESP-derived charges far superior to semiempirical methods, while requiring much less computational effort than even the simplest *ab initio* methods. PESP is also applicable to a large numer of atoms in a wide range of chemical environments, including hypervalent species. Unlike semiempirical methods, PESP charges do not require a uniform scaling factor to improve the statistical agreement with the *ab initio* charges.

The second article in this series will address the applications of the PESP approach to regions of high ESP (e.g., within the van der Waals envelope of the molecule). Future articles will deal with the extension of the method to other atoms. Preliminary parameters are already available for Na^+ and Mg^{++} and parametrizations are planned for K^+, Ca^{++}, and Zn^{++}.

Acknowledgments

The author thanks the Swiss Center for Scientific Computing and Dr. Djordje Maric for a generous grant of CPU time required for the calculation of the *ab initio* ESP-derived charges at the MP2 level.

References

1. E. Scrocco and J. Tomasi, *Adv. Quantum Chem.*, **11**, 115 (1978).
2. P. Politzer and J. S. Murray, In *Reviews of Computational Chemistry*, vol. 2, K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, chap. 7.
3. P. Politzer and D. G. Truhlar, Eds., *Chemical Applications of Atomic and Molecular Electrostatic Potentials*, Plenum, New York, 1981.
4. G. Náray–Szabó, *Chem. Design Automation News*, **8**, 43 (1993).
5. J. G. Ángyán and G. Náray–Szabó, In *Theoretical Models of Chemical Bonding, Part 4, Theoretical Treatment of Large Molecules and Their Interadctions*, Z. B. Maksic, Ed., Springer–Verlag, Berlin, 1991, p. 1.
6. J. E. Douglas and P. A. Kollman, *J. Am. Chem. Soc.*, **102**, 4295 (1980).
7. G. Pepe, D. Siri, and J.-P. Reboul, *J. Mol. Struct. Theochem.*, **14**, 289 (1981).
8. H. Weinstein, R. Osman, and J. P. Green, In *Computer Assisted Drug Design*, E. C. Olson and R. E. Christoffersen, Eds., ACS Symposium Series 112, American Chemical Society, Washington, D.C., 1979.
9. F. A. Momany, *J. Phys. Chem.*, **85**, 592 (1978).
10. S. R. Cox and D. E. Williams, *J. Comput. Chem.*, **2**, 304 (1981).
11. U. C. Singh and P. A. Kollman, *J. Comput. Chem.*, **5**, 129 (1984).
12. L. M. Chirlian and M. M. Francl, *J. Comput. Chem.*, **8**, 894 (1987).
13. C. M. Breneman and K. B. Wiberg, *J. Comput. Chem.*, **11**, 361 (1990).
14. C. I. Bayly, P. Cieplak, W. D. Cornell, and P. A. Kollman, *J. Phys. Chem.*, **97**, 10269 (1993).
15. M. L. Connolly, *J. Appl. Crystallogr.*, **16**, 540 (1983).
16. B. H. Besler, K. M. Merz, Jr., and P. A. Kollman, *J. Comput. Chem.*, **11**, 431 (1990).

17. K. M. Merz, Jr., *J. Comput. Chem.*, **13**, 749 (1992).
18. (a) C. Giessner-Prettre and A. Pullman, *Theor. Chim. Acta*, **25**, 83 (1972); (b) C. Giessner−Prettre, and A. Pullman, *Theor. Chim. Acta*, **37**, 335 (1975).
19. G. G. Ferenczy, C. A. Reynolds, and W. G. Richards, *J. Comput. Chem.*, **11**, 159 (1990).
20. C. A. Reynolds, G. G. Ferenczy, and W. G. Richards, *J. Mol. Struct. Theochem.*, **256**, 249 (1990).
21. P. L. Cummins and J. E. Gready, *Chem. Phys. Lett.*, **174**, 355 (1990).
22. J. J. P. Stewart, *J. Comput. Aided Mol. Design*, **4**, 1 (1990).
23. G. P. Ford and B. Wang, *J. Comput. Chem.*, **14**, 1101 (1993).
24. B. Wang and G. P. Ford, *J. Comput. Chem.*, **15**, 200 (1994).
25. J. E. Wampler, *J. Chem. Inf. Comput. Sci.*, **35**, 617 (1995).
26. T. A. Halgren and W. N. Lipscomb, *J. Chem. Phys.*, **58**, 1569 (1973).
27. T. A. Halgren and W. N. Lipscomb, *Proc. Natl. Acad. Sci. USA*, **69**, 662 (1972).
28. D. S. Marynick and W. N. Lipscomb, *Proc. Natl. Acad. Sci. USA*, **79**, 1341 (1982).
29. A. Deresckei−Kovacs and D. S. Marynick, *Int. J. Quantum Chem.*, **58**, 193 (1996).
30. A. Deresckei−Kovacs, D. E. Woon, and D. S. Marynick, *Int. J. Quantum Chem.*, **61**, 67 (1997).
31. A. Derecskei−Kovacs and D. S. Marynick, *Int. J. Quantum Chem.*, to appear.
33. P.-O. Löwdin, *J. Chem. Phys.*, **18**, 365 (1950).
33. See for instance: (a) F. U. Axe and D. S. Marynick, *Organometallics*, **5**, 572 (1987); (b) F. U. Axe and D. S. Marynick, *J. Am. Chem. Soc.*, **110**, 3728 (1988); (c) L. M. Hansen and D. S. Marynick, *J. Am. Chem. Soc.*, **110**, 2358 (1988); (d) C. A. Jolly and D. S. Marynick, *J. Am. Chem. Soc.*, **111**, 7968 (1989); (e) L. M. Hansen and D. S. Marynick, *Organometallics*, **8**, 2173 (1989); (f) L. M. Hansen and D. S. Marynick, *Inorg. Chem.*, **29**, 2482 (1990); (g) M. Lawless and D. S. Marynick, *Inorg. Chem.*, **30**, 3547 (1991); (h) J. R. Rogers, O Kown, and D. S. Marynick, *Organometallics*, **10**, 2817 (1991); (i) M. Lawless and D. S. Marynick, *J. Am. Chem. Soc.*, **113**, 7513 (1991); (j) P. N. V. Pavan Kumar and D. S. Marynick, *Inorg. Chem.*, **32**, 1857 (1993); (k) J. R. Rogers, T. P. S. Wagner, and D. S. Marynick, *Inorg. Chem.*, **33**, 3104 (1994); (l) J. R. Rogers, C. K. Johnson, and D. S. Marynick, *Inorg. Chem.*, **33**, 4566 (1994).
34. M. J. Frisch, G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson, M. A. Robb, J. R. Cheeseman, T. Keith, G. A. Peterson, J. A. Montgomery, K. Raghavachari, M. A. Al-Laham, V. G. Zakrezewski, J. V. Ortiz, J. B. Foresman, C. Y. Peng, P. Y. Ayala, W. Chen, M. W. Wong, J. L. Andres, E. S. Replogle, R. Gomperts, R. L. Martin, D. J. Fox, J. S. Binkley, D. J. Defrees, J. Baker, J. P. Stewart, M. Head−Gordon, C. Gonzalez, and J. A. Pople, *Gaussian 94*, revision D.2, Gaussian, Inc., Pittsburgh, PA, 1995.
35. W. J. Hehre, L. Radom, R. v. R. Schleyer, and J. A. Pople, *Ab Initio Molecular Orbital Theory*, Wiley, New York, 1986.
36. Original program: M. Dupuis, D. Spangler, and J. J. Wendoloski, National Resource for Computations in Chemsitry, Software Catalog, University of California, Berkeley, CA, 1980, Program QG01. This version: M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. J. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, *J. Comput. Chem.*, **14**, 1347 (1993).